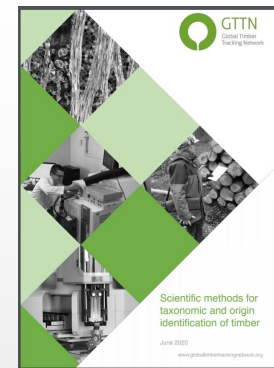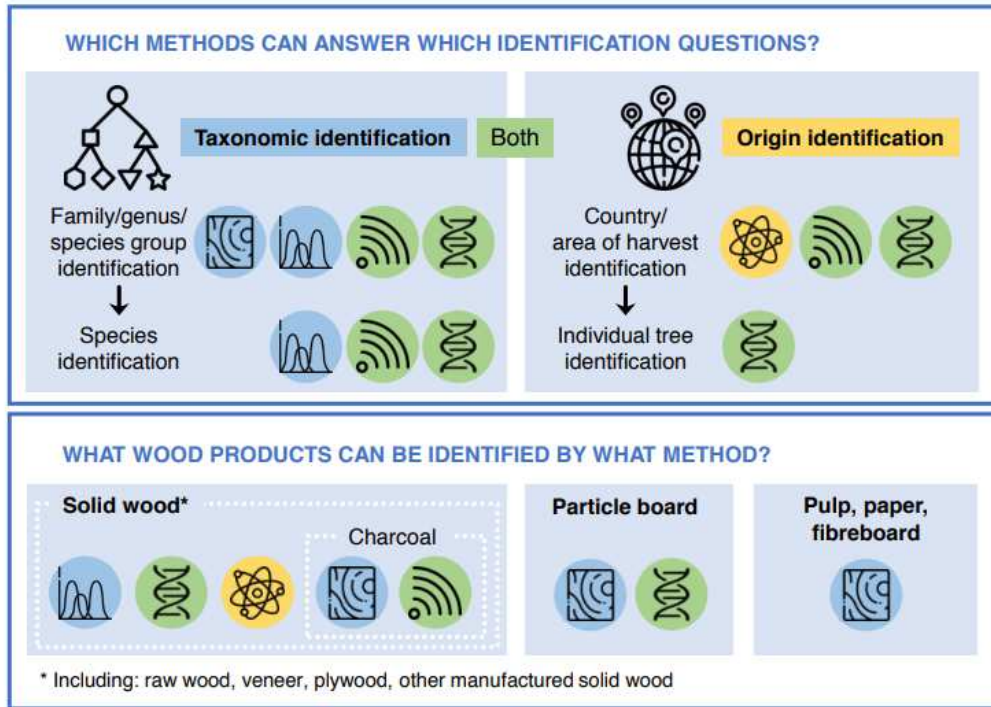# Tracing the geographical origin of timber with genetic analysis

Céline Blanc-Jolivet, Malte Mader, Niels Müller, Bernd Degen
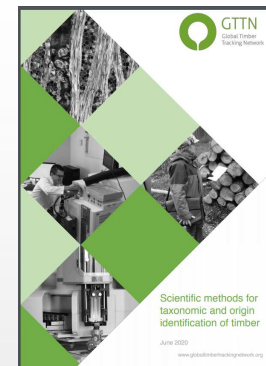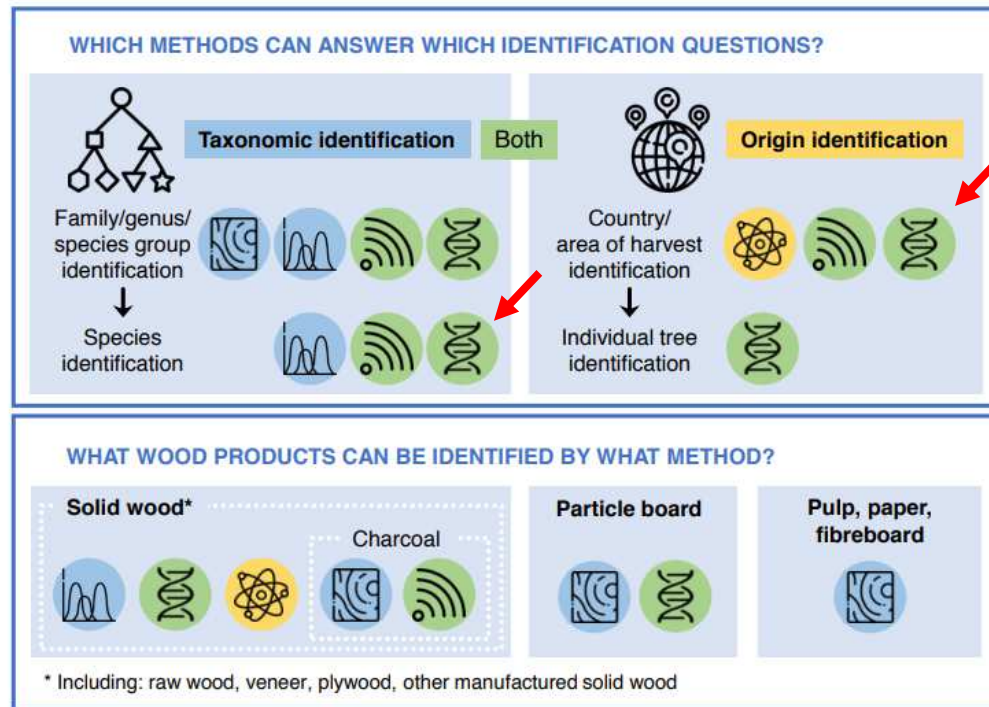Thünen Institute of Forest Genetics, Grosshansdorf, Germany

WoodD+ 15.06.2023

Which methods are available?

The question is… the question to answer !

Céline Blanc-Jolivet
WooD+

THÜNEN

Which methods are available?

The question is… the question to answer !

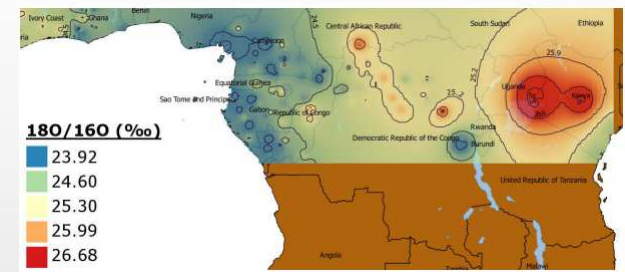**Céline Blanc-Jolivet**
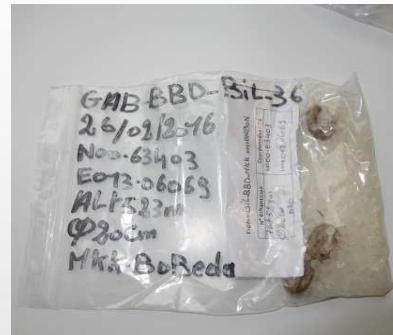WooD+

THÜNEN

Development of reference data

Dried Leaf/
cambium: **DNA**

Wood: **other methods**
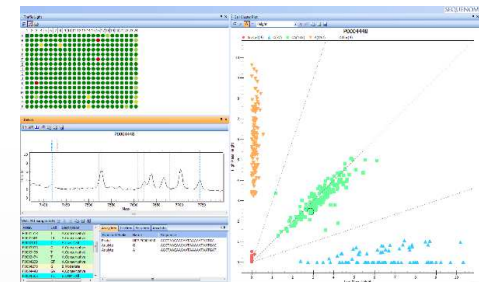
Reference data

Georeferenced reference trees with known species

CITESwoodID

18O/16O (‰)
- 23.92
- 24.60
- 25.30
- 25.99
- 26.68

Céline Blanc-Jolivet
WooD+

THÜNEN
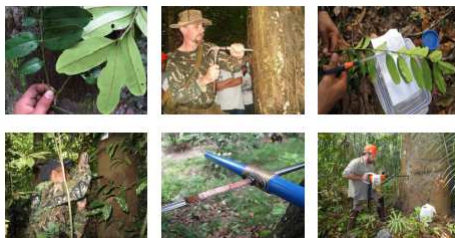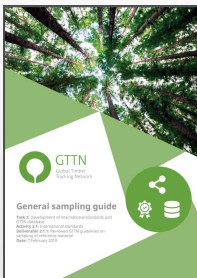
Genetic reference data ?

- **Important: the question is... the question to answer**

- **Geographical origin: which scale?**

  - Country?

  - Region?

  - Forest concession/stand?

- **The scale chosen defines the spatial resolution of the sampling and the analysis**

Céline Blanc-Jolivet
WooD+

THÜNEN

How do I get the appropriate genetic information?

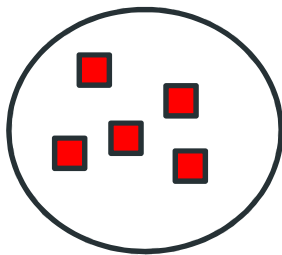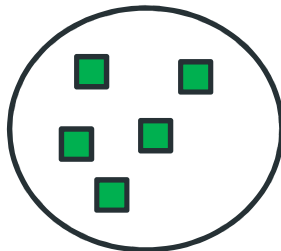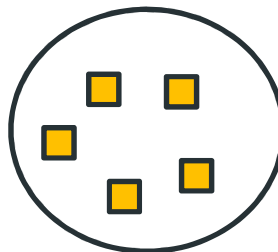- **The ideal case: the genotypes are "region specific"**

Country/Region 1   Country/Region 2   Country/Region 3   Country/Region 4

- **This would likely happens with random genetic information:**
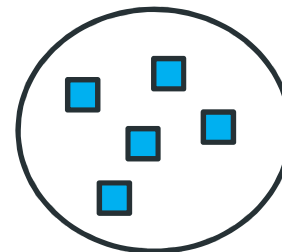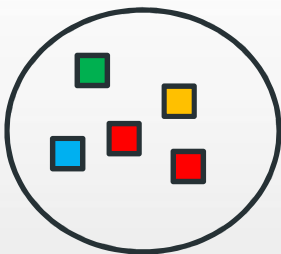

Country/Region 1   Country/Region 2   Country/Region 3   Country/Region 4
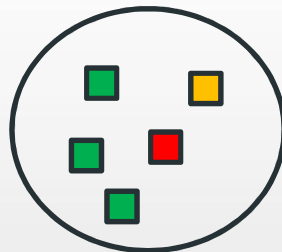
Céline Blanc-Jolivet
WooD+

THÜNEN
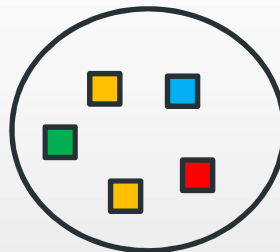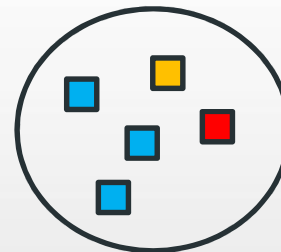
How do I get the appropriate genetic information?

- **How can I find the interesting loci among several millions?**

NGS sequencing of a subset of individuals

selection of loci for cost-effective screening

Céline Blanc-Jolivet
WooD+

THÜNEN

Analysis of NGS data

Raw data: millions of reads

Mapping, variant calling, QC

Genotypes: all loci

?

Genotypes: selected loci

Céline Blanc-Jolivet
WooD+

THÜNEN

Locus selection: example ddRADseq data Beech

5,464 loci, 54 individuals

Missingness <10% ?: **No;** Non-informative loci (MAF<1%) ?: **Yes 2,476**

2,988 loci, 54 individuals

clustering

population genetics parameters

Geographically informative loci (PCA, DAPC): **63** (Europe), **39** (Germany)

Genetic diversity (Hs, Ar), HWE (Fis) within groups: **325** (countries), **305** (Länder Germany)

559 loci

Céline Blanc-Jolivet
WooD+

THÜNEN

Locus selection: example ddRADseq data Beech

Is loss of information after marker selection acceptable?
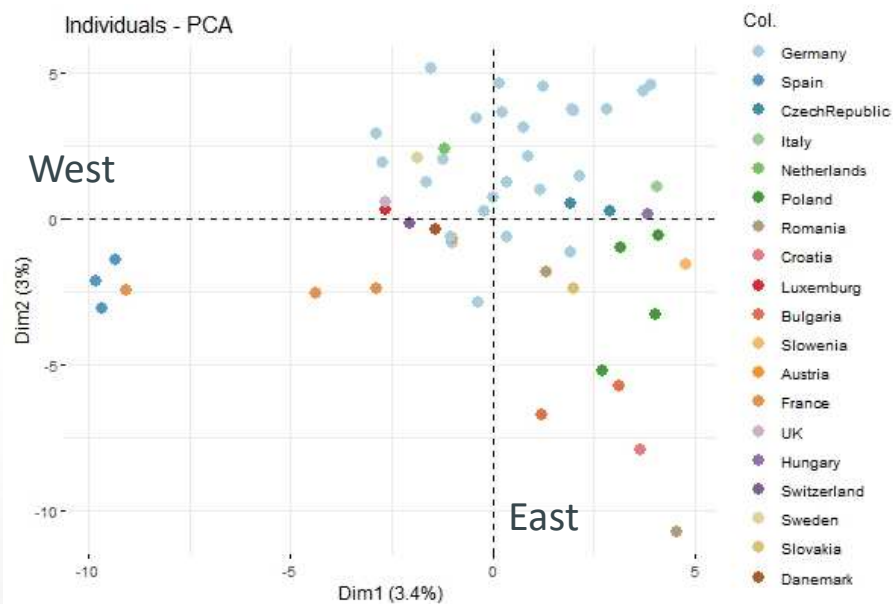
5,464 loci, 54 individuals

559 loci, 54 individuals

Céline Blanc-Jolivet
WooD+

THÜNEN
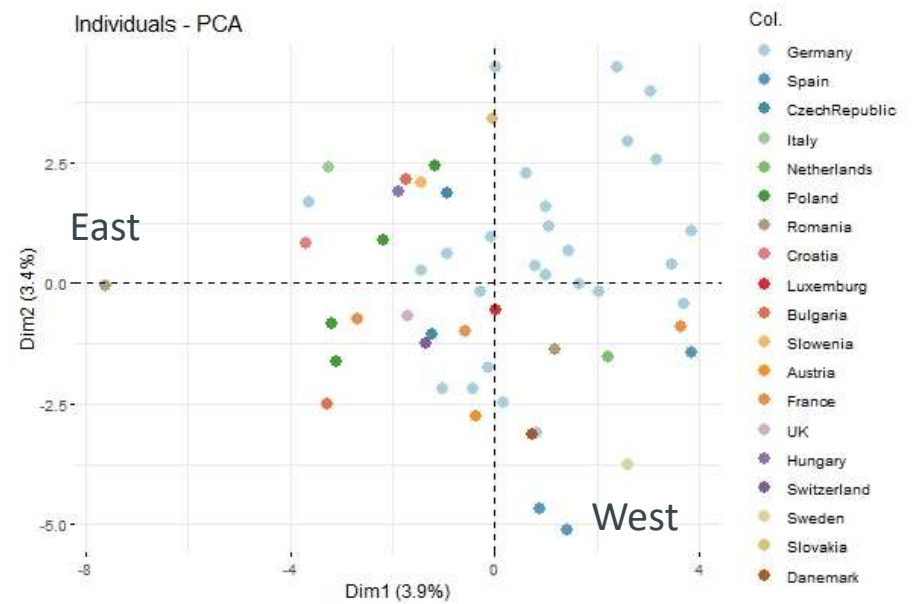
Reference data beech

- **2,000 individuals**

- **543 loci (SeqSNP method)**

- **Cleaning of data: 1,969 individuals @ 507 loci**

    - 31 individuals with >10 % missingness or duplicates

    - 10 loci with > 20 % missingness

    - 7 non-informative loci with rare alleles (MAF < 1%)

    - 19 loci with heterozygote excess or deficit

Céline Blanc-Jolivet
WooD+

THÜNEN

Reference data beech
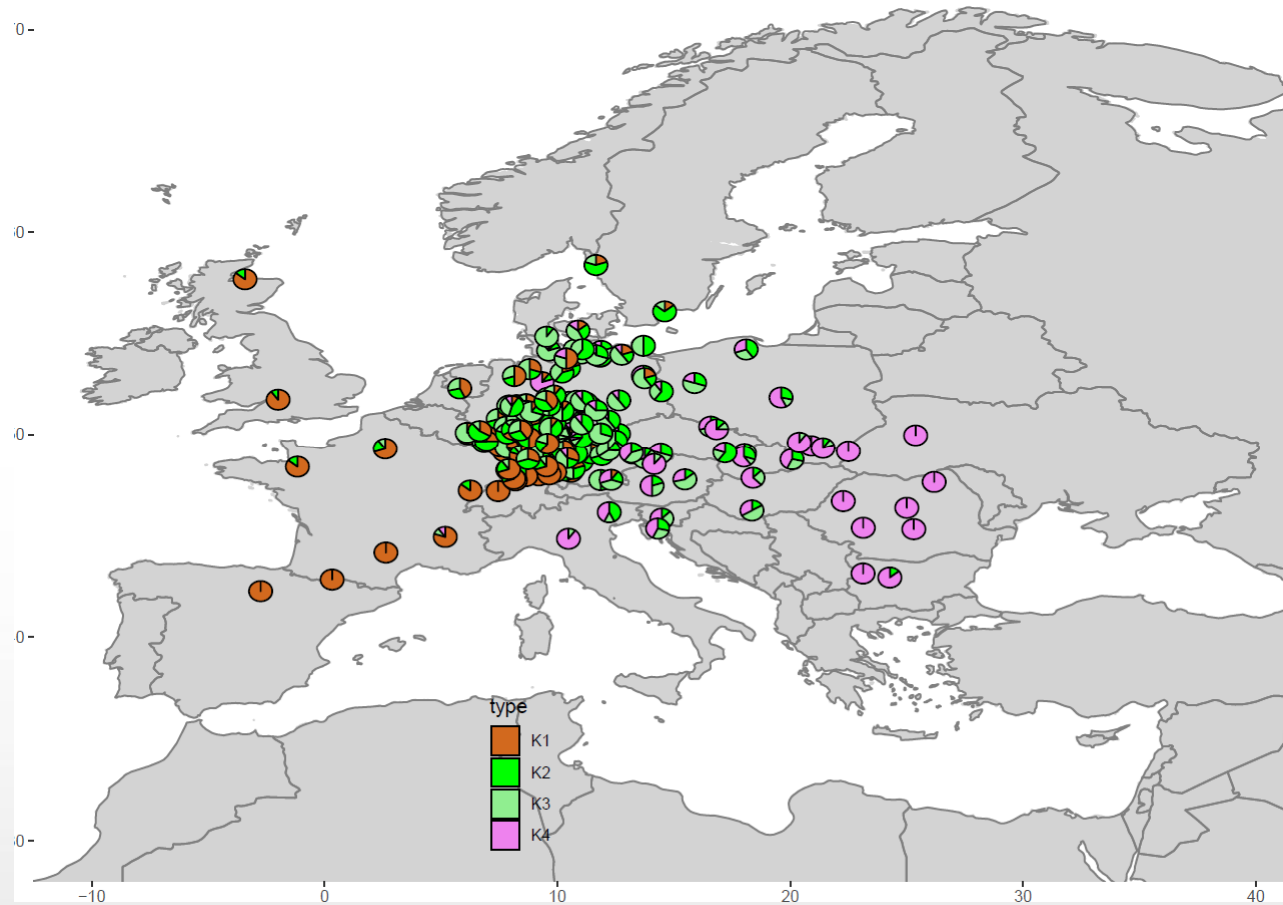
3 genetic groups

DAPC analysis

Céline Blanc-Jolivet
WooD+

THÜNEN

Whole-genome sequencing: higher resolution

860 individuals, 3.6 millions variants

DAPC analysis, 6 main groups

Top variants on first 3 axis: **92**

Céline Blanc-Jolivet
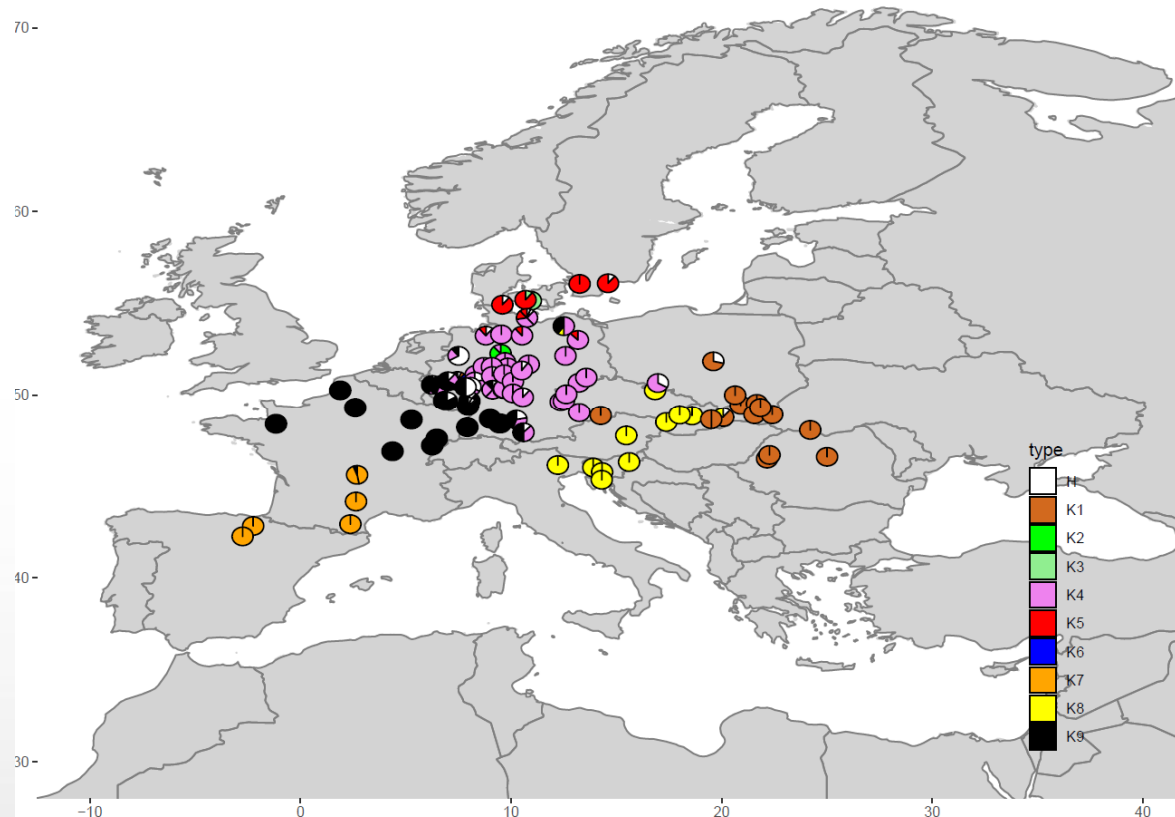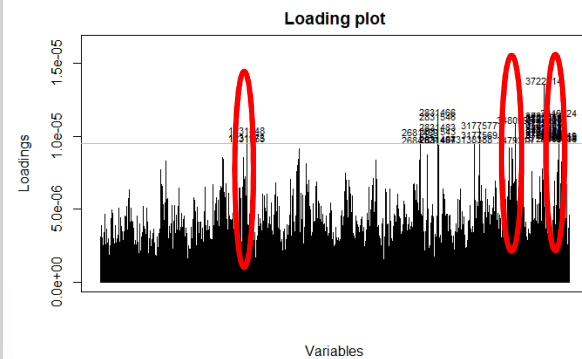WooD+

THÜNEN

Whole-genome sequencing: higher resolution

- **From 92 variants 28 "non redundant" variants selected**

- **Performance of the marker set tested with self-assignment**

Group	Correct	Total	% correct
K1	107	127	84.3
K4	116	307	37.8
K5	39	45	86.7
K7	42	51	82.4
K8	46	86	53.5
K9	108	171	63.2
Total	458	787	58.2

Céline Blanc-Jolivet
WooD+

THÜNEN
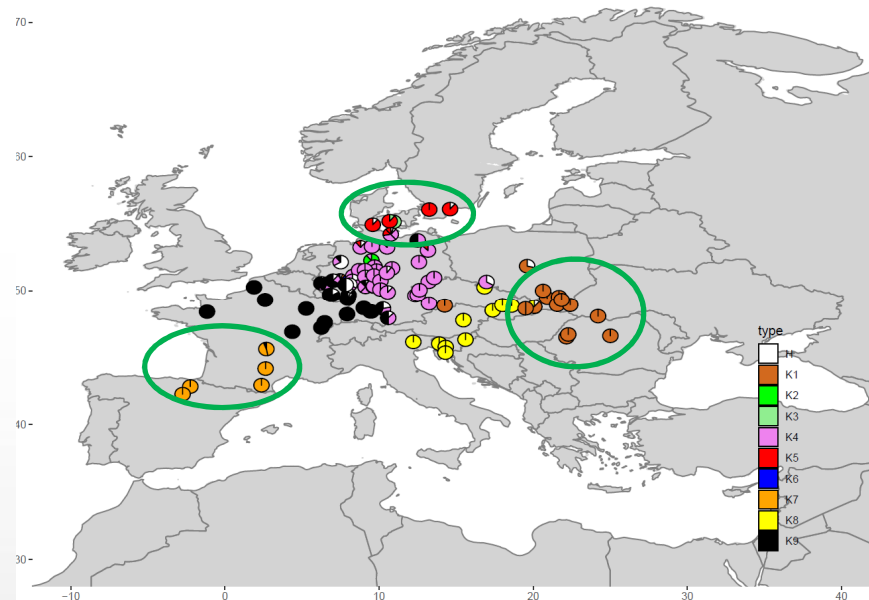
Analysis of SGS disentangle complex situations

- **Tropical species: taxonomy unclear**

- **Species declarations on timber mostly concerns the main well-known species, not all**

  - Cumaru: *Dipteryx odorata*

  - Ipe: *Handroanthus serratifolius/impetiginosus*

  - Jatoba: *Hymenaea courbaril*

- **Set of molecular markers can in some genera both identify species and origin**

Céline Blanc-Jolivet
WooD+

THÜNEN

Testing species and origin with the same method

Easy case: species have different distribution ranges

Larix sp.

Céline Blanc-Jolivet
WooD+

THÜNEN
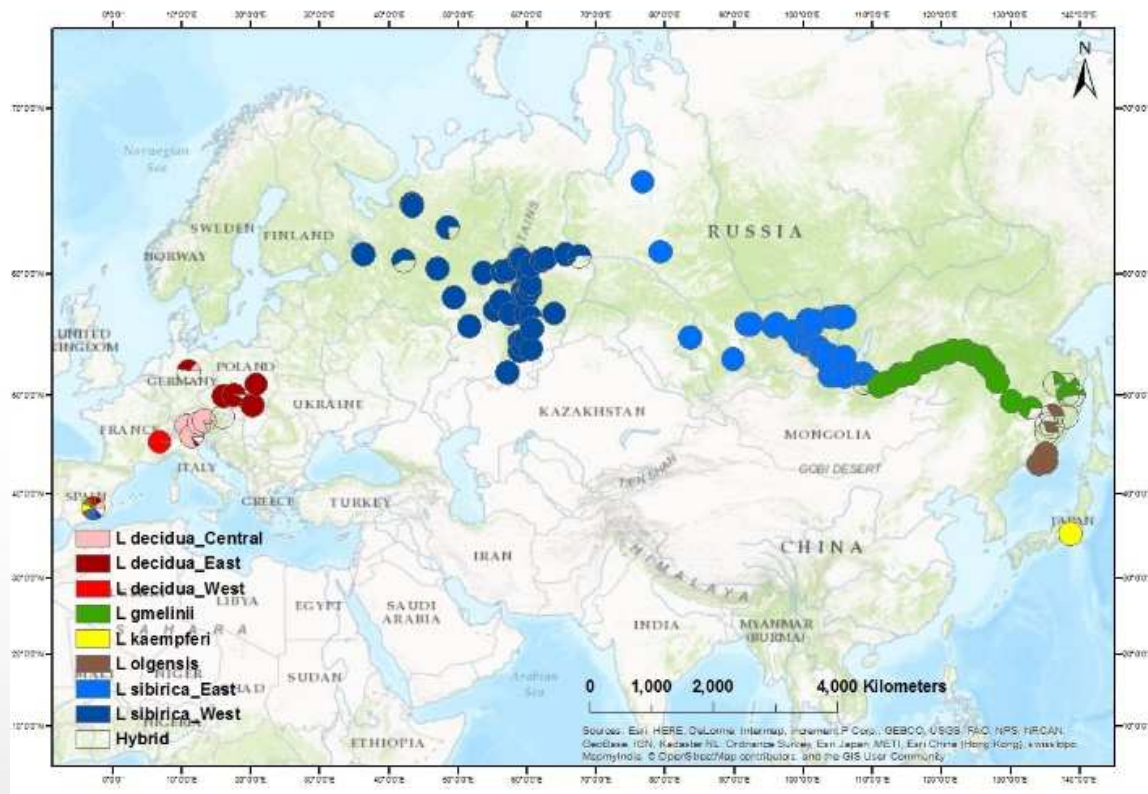
Testing species and origin with the same method

Easy case: species have different distribution range, BUT species could be only identified after analysis
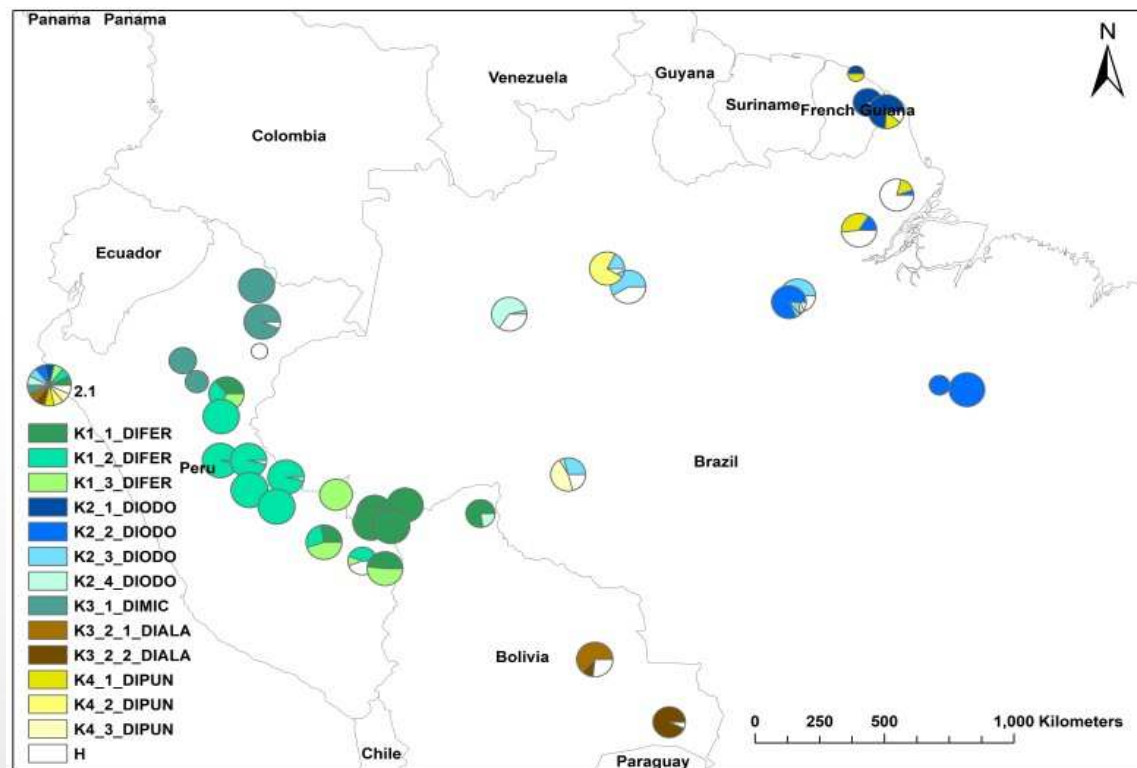
Dypterix sp.
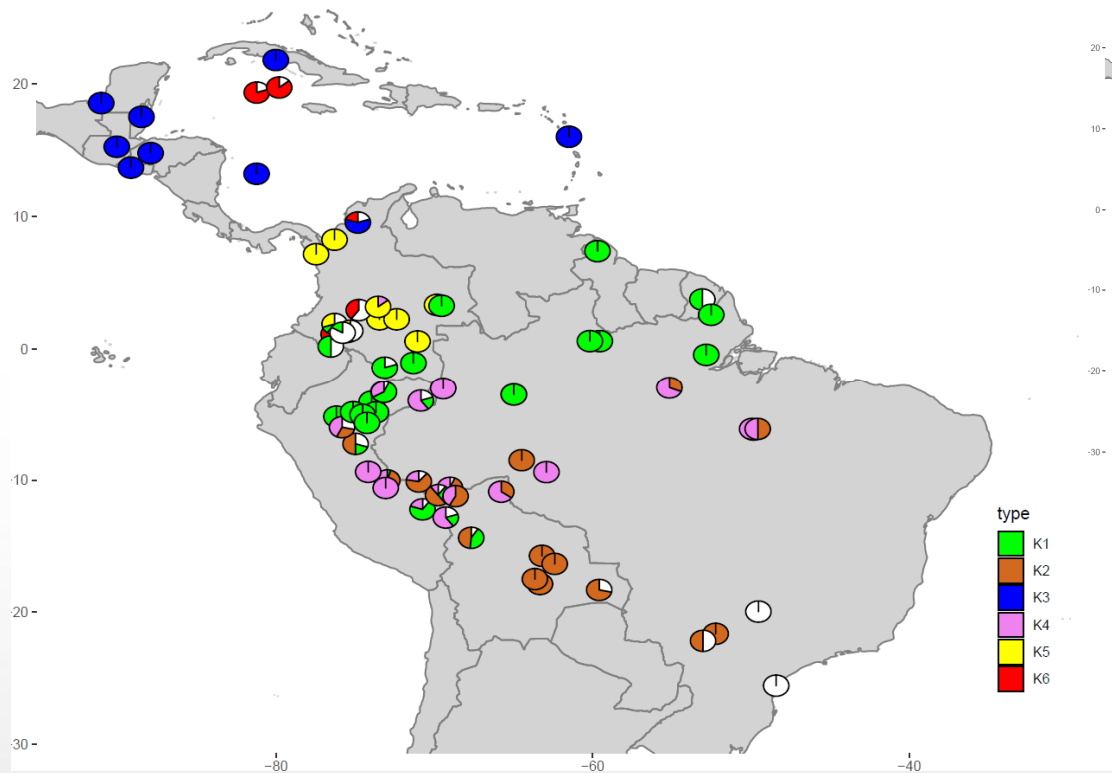
Céline Blanc-Jolivet
WooD+

THÜNEN

Testing species and origin with the same method

More difficult: some species are sympatric in some regions and plastidial DNA gives no species info

Cedrela odorata, C. fissilis (K2), nuclear DNA only
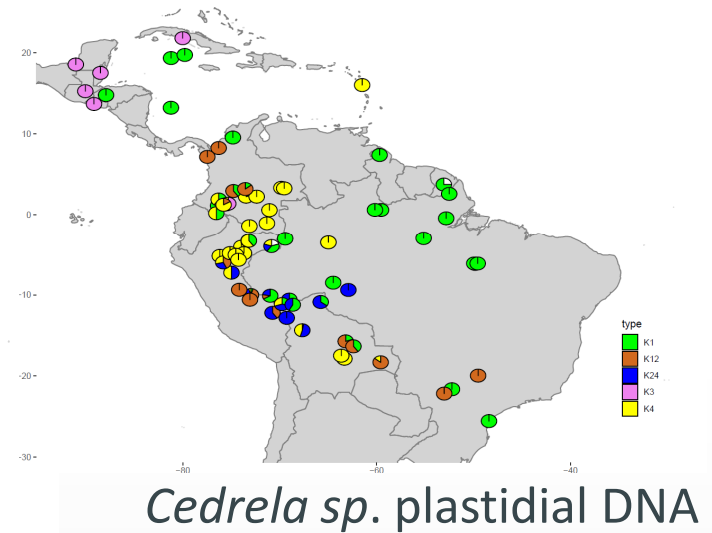
Cedrela sp. plastidial DNA

Céline Blanc-Jolivet
WooD+

THÜNEN

Challenging case: species are sympatric AND unclear taxonomy
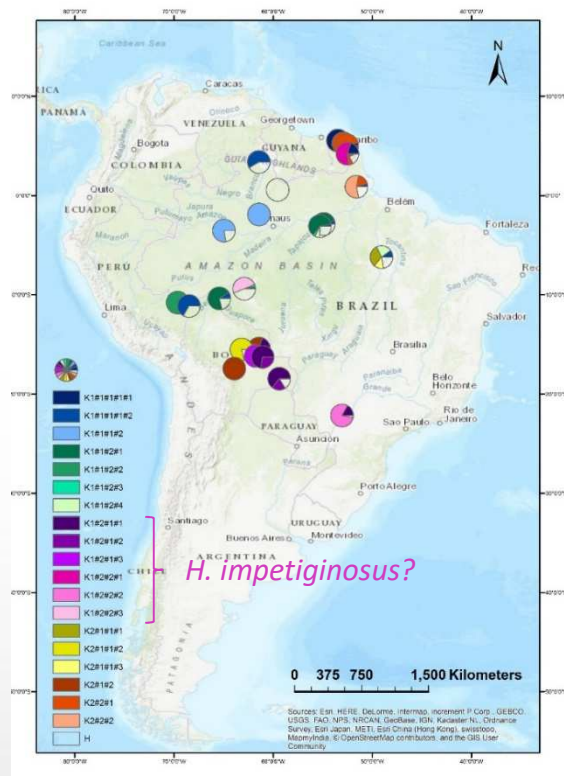
Handroanthus sp. (Nuclear + Plastid)
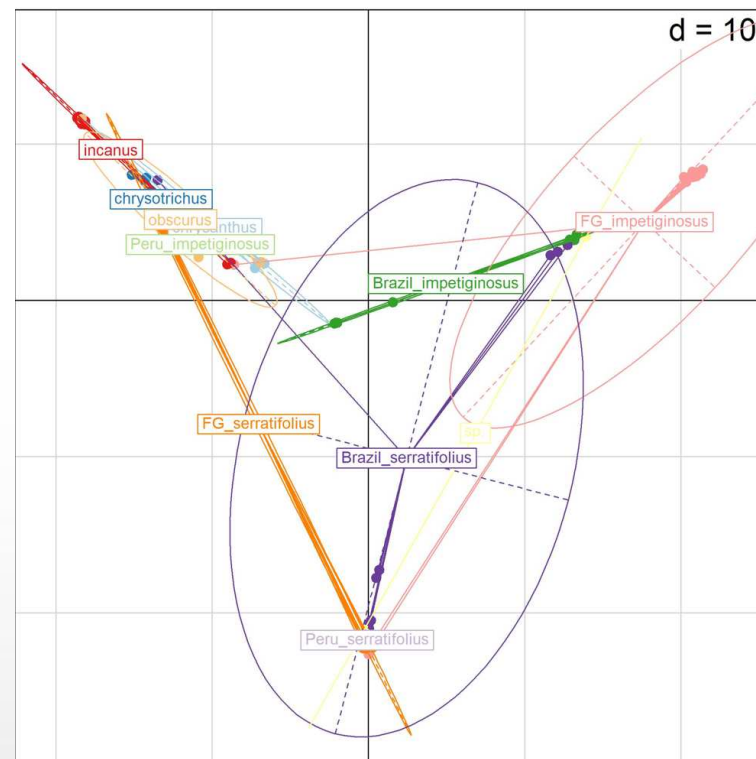
Handroanthus sp.: species ID loci?

Céline Blanc-Jolivet
WooD+

THÜNEN

Testing species and origin with the same method

Cryptic species?
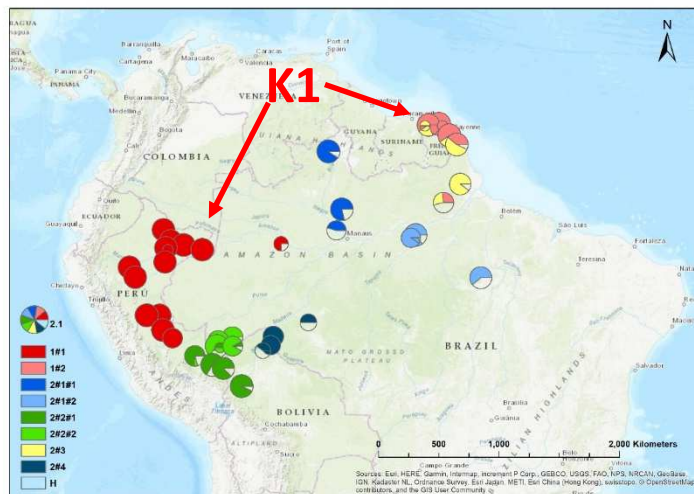
Jacaranda copaia (nuclear)

Lophira alata (Nuclear)

Nauclea diderrichii (Nuclear)

Céline Blanc-Jolivet
WooD+

THÜNEN
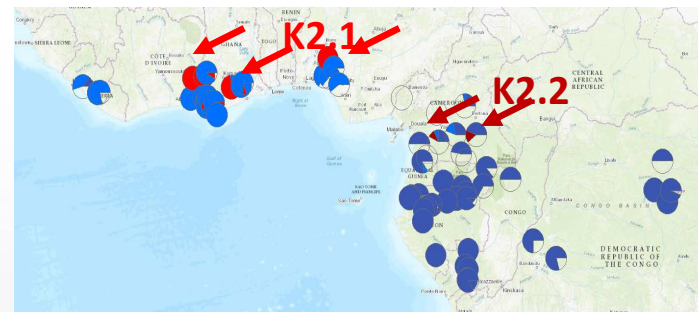
The methods are species-specific… but

- Each species/species complex has a different spatial pattern

=> separate reference data development

=> new geographical groups are identified

=> marker sets are specific

- Some geographical patterns are common among species

=> evaluate chances of success!!

Céline Blanc-Jolivet
WooD+

THÜNEN

Common patterns among species

Entandrophragma utile (Nuclear+ Plastid)

Cylicodiscus gabunensis (Nuclear)

Nauclea diderrichii (Nuclear)

Céline Blanc-Jolivet
WooD+

THÜNEN
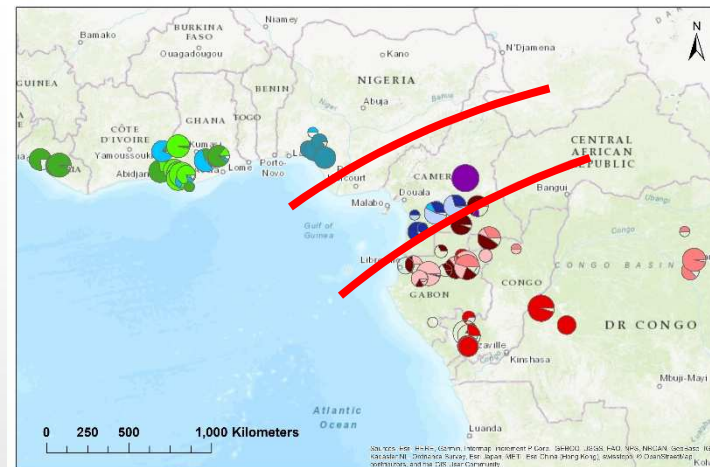
Which steps are necessary to apply the methods?

Testing the reliability of the method (for origin ID)

- ## Statistical power
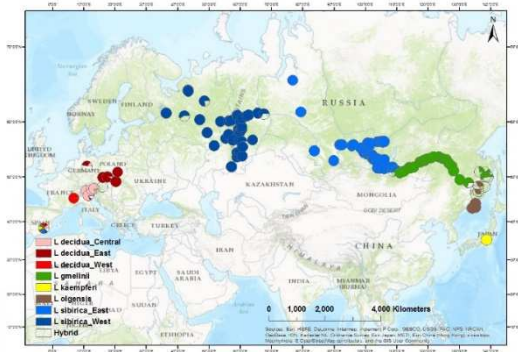

```
##########################################
###      Results of Self-Assignement       ###
##########################################

Group Size                                  = 1
Number of Tests                             = 1687
Method of assignment                        = Bayesian Approach
Minimum number of bi-parental loci requested = 1


Population              Sample Size    Tested ind/groups    % correct assigned
-------------------------------------------------------------------------------
L decidua_West              59              59                    46
L decidua_Central           85              85                    73
L decidua_East              61              61                    49
L sibirica_West            626             626                    98
L sibirica_East            503             503                    97
L gmelinii                 250             250                    83
L olgensis                  68              68                    93
L kaempferi                 35              35                   100


Total                                                             90
Mean                                                              80
```
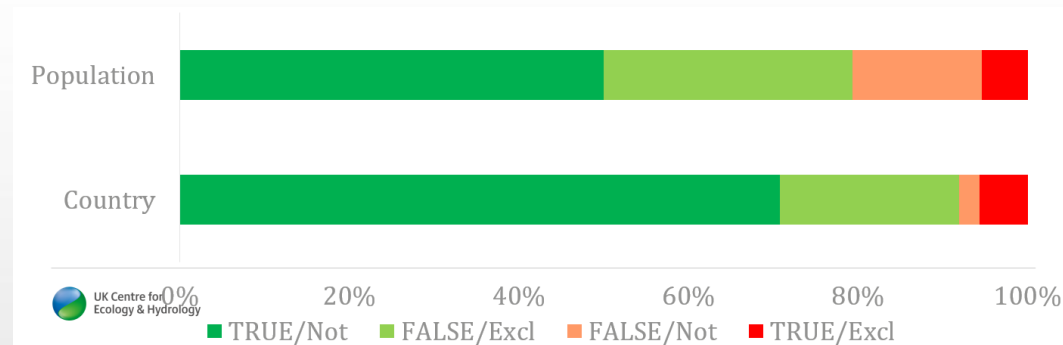
- ## Blind testing

Céline Blanc-Jolivet
WooD+

THÜNEN

Thanks !

Hilke Schroeder

Björn Klinkhammer

Stefanie Palczewski

celine.blanc-jolivet@thuenen.de

THÜNEN